

Bard

Bard College
Bard Digital Commons

Senior Projects Spring 2020

Bard Undergraduate Senior Projects

Spring 2020

Friend or Foe? Outstanding Peer's Unexpected Impact on Test Taking Ability

Woanjun Lee
Bard College, wl1664@bard.edu

Follow this and additional works at: https://digitalcommons.bard.edu/senproj_s2020



Part of the [Education Commons](#), and the [Social and Behavioral Sciences Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Lee, Woanjun, "Friend or Foe? Outstanding Peer's Unexpected Impact on Test Taking Ability" (2020).
Senior Projects Spring 2020. 246.
https://digitalcommons.bard.edu/senproj_s2020/246

This Open Access work is protected by copyright and/or related rights. It has been provided to you by Bard College's Stevenson Library with permission from the rights-holder(s). You are free to use this work in any way that is permitted by the copyright and related rights. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself. For more information, please contact digitalcommons@bard.edu.

Bard

Friend or Foe?

Outstanding Peer's Unexpected Impact on Test Taking Ability

Senior Project Submitted to

The Division of Science, Mathematics, and Computing

of Bard College

By

Woanjun Lee

Annandale-on-Hudson, New York

May 2020

Abstract

Many studies have discussed methods of improving individuals' performance in a variety of fields. Furthermore, studies also have observed how social interactions may have an impact, whether positive or negative, on an individual's performance. The current study aims to expand on this thread and take a deeper scope on the immediate impact on academic test-taking ability of the presence of a confident colleague. Additionally, the current study explores whether the change of performance, if any, impacted by the stimuli is exaggerated depending on the subject's sex or extraversion. The current study assigned 14 Bard undergrad students to the control group, and 7 Bard undergrad students to the experimental group. All participants were asked to take two short mathematics tests consecutively, one easy and one difficult, with no time limit. The experimental group had a confident confederate acting as another participant that worked on the same task next to the subject and was asked to finish the test drastically earlier than average. I hypothesized that participants in the experimental group would perform the task significantly faster than the control groups in both tests, but would show impaired accuracy in both tests. I also hypothesized that there will be a difference in effect depending on the participant's sex and extraversion. The results showed that there was no difference in test scores between groups, but participants in experimental condition performed faster in the difficult exam. Additionally, difference in sex did not affect the effect of the confederate's presence in participant's score or time, but extraversion did affect in terms of score.

Friend or Foe? Outstanding Peer's Unexpected Impact on Test Taking Ability

It is no surprise that people like to see an improvement in their performances. Regardless of the domain, such as sports, academics, or music, increasing performance is many individuals' goals throughout their life or career as it could provide a general sense of positivity. Such positive emotion can come from a sensation of acceptance, achievement, pride, or better opportunities. As a result, many studies have long been interested in finding efficient and effective ways to improve one's task performance. In particular, academic education in the United States (U.S.) is a key component to young people's lives, as many states in the U.S. require individuals from the ages of 6–18 to be a full-time student until they graduate high school. As a result, improving a student's academic performance is an interest of students, parents, teachers and researchers.

Regarding academic performance, many longitudinal studies illustrated that having a certain mindset or support group around them can improve performance, or in this case, grades. For example, it is suggested that different social support groups have a crucial impact on adolescent students' stronger mastery goals, performance-approach goals, performance-avoidance goals, and test anxiety (Song & Bong, 2015).

Studies have explored and illustrated that certain styles of teaching and counseling predict more positive results on improving students' academic performances. For example, a classroom environment that reflects a common mid-20th century style, where teaching and learning is done mutually by the students and faculty, show significantly greater student's comfort, confidence as well as engagement level within the class (Wright, 2020). A therapy style that led the student to focus and explore the roots of their lack of success also seems to

significantly benefit students' improvement in grades (Esposito, Karterud, Freda, 2019).

Attribution therapy, providing students with a mere possibility of success, or simply a model they can look on, also seems to be a sufficient means of aiding students in their academic performances. In Wilson & Linville (1982)'s study, college freshmen who saw a video of an upperclassmen in their school talking about how their grades improved than the time in their freshmen year, improved their GPAs significantly more than those who did not.

It becomes more clear that progressively changing a college student's attitude towards realistic and constructive ways can benefit students greatly. Students who have such attitudes show greater academic success, as well as better mental and physical health. (Chemers, Hu & Garcia, 2001). Additionally, certain independent ambitions and high maturity of students seems to improve grade performance (Gelso, Osterhouse, & Bodden, 1971).

Speaking of a student's attitude, Self-determination theory (SDT) must be discussed. First suggested by Deci and Ryan (1985), SDT distinguishes motivations that are intrinsic (autonomous) or extrinsic (controlled) based on an individual's personality within social contexts. Those who are intrinsically motivated to their work, tend to be, not only be more engaged in their work, but also were less likely to quit during poor performances, and eventually show more promising results than those who were not.

Obviously, such in-depth scope on factors that contribute to a student's academic outcome in the long run is indubitably beneficial for students growth. However, many would like to see a positive change happen fast. Variables of the longitudinal study, such as certain attitudes, goals, or roots of motivation are not something that can be easily changed in a quick amount of

time. Furthermore, not everyone in the academic field has the luxury of a trustworthy and skilled counselor or support group.

Therefore, it should also be important to know if there's any way to improve student's performance in the short term. Luckily, numerous studies have shown the possibility of directly impacting one's performance through the addition of certain stimulus, though not limited to academic performance. Regardless, further exploring these stimuli that show immediate effect on an individual's performance may hint at better ways to improve students' chances of academic achievement and success.

A1. Social Impact by Peers and Social Interaction. First and foremost, overall, it is unequivocally clear that performers are greatly affected by social interactions during a task. The relationship between social interaction and performance is noteworthy as many tasks individuals perform are within a social setting. For instance, academic tests are usually taken amongst peers, and athletic and musical performances are often done in front of an audience. Many studies have shown the effects of social facilitation in both humans and animals, where individuals often perform significantly better when they are doing the task with other people if the task was easy (i.e., the fishing reel task studied by Triplett, 1898).

Some have shown the opposite effect, where participants' performances were noticeably impaired based on social observation (Michaels et al. 1982). Interestingly, this contradictory result is paralleled in cockroaches, where cockroaches surrounded by other cockroaches were able to go through mazes significantly more quickly when the maze was simple, but struggled when the maze was difficult (Zajonc, Heingartner, Herman, 1969).

Additionally, skill level and comfort also seem to be variables that affect the correlation between performance and social interactions. In the Pool Hall study (Michaels et al. 1982), pool players of different skill levels (one being above the average skill level and another being below the average skill level) showed different effects when an audience was present. Skilled pool players showed improvement in performance in the presence of an audience, whereas unskilled pool players showed impairment in performance in the presence of an audience.

Other studies illustrate that there are many more complex interactions that affect individual performance based on social settings. For instance, Social Loafing Theory (Kravitz and Martin, 1986; Ringelmann, 1913) suggests that individual performance decreases when that individual is in a group where labor is pooled. That is, when individuals are working in a group together, they are likely to put in less effort than they would if they were working on the same task alone (Ringelmann, 1914).

Furthermore, Spencer et al. (Spencer, Logel, Davies, 2016) suggests that performance is also affected beyond the immediate social interaction that is present, and also based on social biases and stereotypes. Specifically, Spencer suggests the idea of Stereotype Threat, where individuals are under a perpetual state of pressure to perform better to not misrepresent groups of which they are a member, following exposure to inaccurate stereotypes about their group. For example, if there is a cultural stereotype present that women are likely to perform worse than men in mathematical tasks, women who are performing the task are likely to have their performance impaired and perform in an inferior manner to a man that has the same qualification as them, due to the fact that they are under pressure (Spencer, Steele, & Quinn, 1999). Based on such promising findings of effect of social interaction on task performance, the current study

hopes to see an effect of task performance of students when they are under *peer pressure*, where students are not only performing the task with another peer (without cooperating with each other), but also have the peer finish the task extremely quickly and confidently.

A2. Social Impact of Expectations. Previous studies on the roles of expectations of parents, teachers, and students themselves show interesting results. Many studies illustrate a strong correlation between scholar/model roles' expectations and the students academic outcome, that is, expectations of a teacher predict the academic achievement of children (Kolib & Jussim, 1994); children who are gifted showed negative academic outcomes when their teachers had negative academic expectations to them. Moreover, parallel relationships were illustrated to be present in a study by Doren, Gau, and Lindstrom et al (2012), where positive academic expectations from respective parents predicted positive academic outcomes from their children who had disabilities in learning. Such effects were also parallel in self expectancy as well. Urdan et al. (2007) also hints the contrary where negative parental expectation can be a strong motivation factor that drives students in hopes of disproving societal and parental expectations. (Rubie-Davis, Peterson, Irving et al, 2010)

One important factor to note is the self-fulfilling prophecy, where individuals who believe in themselves and have higher sense of goal will likely to work harder. This expands to parents and teachers, who will likely to provide positive feedback, proper and better guidance and resources to individuals who they believe to have a better chance of academic success. Self-fulfilling prophecy may potentially be a main part of the mechanism of how expectancy from self and environment predicts academic performance.

One of the most famous illustrations of this is, of course, Rosenthal and Jacobson's Pygmalion in the Classroom study (1968). This study showed the high-expectations of the teachers or instructors alone, in the long run, can drastically increase the IQ levels of the students. In early stages of elementary school, some teachers were told that certain students are "growth spurters", who are inclined to greatly succeed in the future. Yet, these students were normal who were merely randomly assigned to the experimental condition. The outcome was remarkable; over time, students who were marked as "growth spurters" almost doubled in IQ gain in six grades in lower groups. Rubovits and Maehr (1971) suggested "interaction quality hypothesis" to serve as a potential explanation. They found that there was no difference in the time or attention that the teacher gives to the students in either condition. Rather, the style of attention and compliments provided varied significantly depending on the student's condition.

Interestingly, recent studies also revealed that beyond teachers' or general educators' expectations towards students, parents' anxiety and fear of a certain subject may predict their children's performances in that area. For instance, there is a direct negative correlation between parents' anxiety level towards math and their young children's early math performance (Shaeffer, Rozek, Berkowitz, Levine & Beilock, 2018). Shaeffer et al. (2018) then further discussed how certain styles of intervention involving friendly and mutual learning of students and parents can break this correlation. This possibility is credited to changes in parents' expectations towards their children's success in math and ultimately shows that parents' expectations can play a significant role in deciding and potentially improving students academic performances (Shaeffer et al. 2018).

However, it is important to understand that the mentioned prior studies were usually longitudinal and far from illustrating the immediate effect of academic performance after intervention of creating high expectancy. Furthermore, unlike the prior study, the students will have no availability to receive positive feedback and guidance from others and themselves as they are performing one task in one sitting. While the pressure to perform well, or to perform to the self-expected threshold is present, and limited availability for self-fulfilling prophecy regarding time and resources, the current study expects the student to perform worse in the presence of the confederate.

B. Impact of Pressure. Nevertheless of the varying studies, the idea of *choking* -performing heavily under one's capability due certain pressure- is omnipresent in diverse fields, such as academics, sports, and jobs. Even with the highest expertise and skills people tend to fail and researchers tried to give an explanation of how and why to such weird, yet common phenomena.

Several researchers contribute working memory as a determining factor of whether or not a high stake situation/environment, (or simply, pressure) can negatively affect students in academic fields. Working memory is part of a short-term memory that people use to apply information during task performances; for instance, recalling a restaurant's address or phone number while driving towards it through traffic is an application of one's working memory. In her book *Choke*, Beldiok (2010) discusses how one's capacity of working-memory is a strong predictor of their academic success, as it directly affects one's ability to perform reading and problem solving.

Ironically, several studies have shown that, in fact, students who have high working-memory are the ones that tend to choke or fail under pressure. Students who have higher working-memories have a pattern of using algorithmic methods (procedural, time-consuming, yet accurate) while students with lower working-memories opt towards certain shortcuts that can lead to flaws (Beldiok, 2010). Because of such differences, students with higher working-memories perform exceptionally in their academic tasks during low-stress conditions. However, the moment pressure rises and creates a high-stress condition, it was observed that many students with high working-memories panicked and opted to the method that the students with lower working-memories traditionally used and performed less than what they were capable of (Beldiok, 2010).

Later studies further explore that capacity of working memories is not the only predictor of susceptibility of choking. Using salivary cortisol as a measure for arousal (pressure), one study illustrated that anxiety is the new addition to the puzzle. Given a novel math examination under pressure (higher salivary cortisol levels), students with high levels of working memories with *high math anxiety* showed significantly worse performance than students with high levels of working memories with *low math anxiety*. In fact, the latter group performed better than usual (Mattarella-Micke, Mateo, Kozak, Foster & Beilock, 2011).

This finding is incredibly important as it suggests that pressure of the circumstances can be converted to an advantage to one's performance depending on their working memory capacity and anxiety. Benny & Banks (2018) further supports and expands this finding. 60 undergraduate students in their study showed lower risk of choking under pressure if they had high working memory and if the task being performed demanded high working memory from the test-takers.

Furthermore, parallel to Matterella-Micke et al.'s former finding (2011), Benny & Banks (2018) also showed students that demonstrated high state anxiety during the task, as well as negative thoughts on the task evaluation predicted higher susceptibility to choking.

Aside from pressure of success in a high-stake, high stress situation, pressure of time is one of the most relevant factors in students' academic lives as their assignments and tests have deadlines and time limits. Though one might presume that giving such time pressure in an academic setting can lead to better results as it may reduce laziness and complacency, some studies show that time pressure may be a significant double-edged sword regarding improving a student's academic performance.

Interestingly, according to McDaniel (1990), auditors in companies showed increase in efficiency but decrease in effectiveness; in this report, "effectiveness" was used as performance meeting professional expectations, and "efficiency" was defined as reaching expected standard while having the lowest cost possible. Results show that limited time, or time pressure in companies show improvement in efficiency of company audits, that is "effectiveness per unit of time", but impair the overall effectiveness produced.

McDaniel mentions a follow up report by Rhode (1978) found that over half the audits performed under time pressure were greatly impaired, performing way below the expectations. However, it is noteworthy that these findings are based on company occupations, individuals are not in an academic setting and it is likely that the individuals are above the age of average college students. The current study is interested in whether the findings of McDaniels are parallel in an academic setting, where students will perform worse on standardized tests, when under time pressure.

Method

Participants

A small group of Bard College undergrad students (14 women, 7 men, $M_{\text{age}} = 20.3$, age range : 18-25) was recruited through posters and flyers that were posted on main facilities of Bard College as well as through the advertisement posted on social media. More participants were further collected through researchers personally recruiting students on campus directly. All participants, regardless of their performance, were promised an equal chance to win a \$30 Amazon gift card after the study. Regardless of their recruitment methods, all participants were falsely informed that the study is about exploring in which environment do students perform best academically. Oblivious to the presence of conditions, each participant who was available at the same time as the confederate was put into the experimental condition and those who were not were put into the control condition. All participants were invited to Preston Lounge at the assigned time without any preparation for data collection.

Materials and Procedure

The Confederate. In order to avoid any additional unknown variables, the current study only used one confederate for all participants assigned to the experimental condition. The Confederate is also a Bard undergrad senior who is a caucasian male with above American average height. The confederate was instructed to not initiate any interaction with the participants, but respond neutrally when the participants did so to him. To assure the confederate's credibility, the confederate was asked to be present at the lab before the participant arrived. Once the participant arrived, the researcher thanked both for coming and provided the informed consent form and written instructions of the task. The written instructions stated that

the participants will be taking a math test (with a calculator and with no time limit), and to not use any electronics and vaguely mentions that they might experience differences in the environment such as abnormal temperature, lighting or sound. Though it was not orchestrated, the confederate made a small comment about the test that seemed to make himself. The current study decided to include it in the procedure as it seemed to give more credibility to the confederate for the participant during the first data collection of the experimental condition.

Both the confederate and the participant were then guided to a room where there are two computers placed about one foot away from each other. Each of the computers have the default microsoft calculator on and internet disabled. When the test was given, participants were reassured that there was no time limit and to “take however long as they need”. Oblivious to the participants, time it took for them to finish the test was measured.

The confederate was instructed to use confident body language throughout the test (frequent nodding, fast typing on the keyboard and fast writing on the paper) and finish and complete the test significantly earlier. (1 minute and 45 seconds for the first test, and 7 minute and 30 seconds for the second test. These times were approximately 15 seconds faster than the average of the time it took for three pilot testers. The pilot testers were the researcher and his family, and generally has above average math experience). After the confederate finished the first test, he was instructed to wait until all participants were done. However, after the completion of the second test, the confederate was asked to “wait in the other room” so the participant did not doubt the confederate, but then was secretly dismissed for their convenience.

The Tests. Because the current study is also interested in how difficulty plays a role in the impact of the confident peer on the participant’s test performance, two tests were designed

with drastically different difficulty. Participant's task consisted of finishing both tests of varying difficulty, but they were graded and timed separately. The first exam used seven simple questions from an 8th grade Algebra textbook and asked to solve variables. The second exam, on the other hand, consisted of seven multiple choice questions directly copied from a mathematical reasoning section from a 2019 GRE test book. Each participant was given the first exam (easy algebra test), and upon its completion, the second one (difficult GRE test) was given.

Each participant received 1 point for getting a question correctly and could score a max of 7 points in either test. Initially, both tests were designed to provide no partial credit and only grade based on the final answer and not the procedure before it (a handful of participants asked if they have to show work during their task). However, one question in the second (difficult) exam gave participants the option to choose up to four answers. Because this unique question allows participants to choose any combination of correct and incorrect choices at the same time, it seemed fair to come up with an alternate way of grading this question. The SAT (pre-2016 version) grading system, where lucky guessing was discouraged by deducting points for incorrect answer, while not deducting any points for unanswered questions inspired the grading criteria for this unique question: of the four options, two were correct; participants received 0.5 points for each correct choice, but also got 0.5 points deducted for each wrong choice. For example, if a participant chose all four options, he or she received a 0 ($1-1=0$) for the question.

Post-task Survey & Manipulation Check. After the completion of the second exam, the participants were given a post-task survey. The survey included five sections : questions regarding basic demographic information, STAI (State-Trait Anxiety Inventory), TIPI (Ten Item Personality Measure), general perception of the first exam, and finally, general perception of the

second exam. Surveys for participants in the experimental condition had an additional sixth section that discussed participants' perception of the confederate.

For all participants, age, class year, and ethnicity were reported for the demographic section. All demographic information was written by participants rather than being chosen from a set of options. The sample was majority White ($n = 15$, 71.4%), with other participants being Asian (14.3%), Biracial (9.5%), and African American (4.8%).

There was no significant difference of demographic information between two groups, except class years. The demographic information included age, ($t[19] = 1.74$, $p = .10$, $M = 20.7$ for control and $M = 19.4$ for experimental) class year, 1 being a freshman and 4 being a senior: ($X^2 = 9.6$, $p = .022$, $N = 21$), gender ($X^2 = 1.71$, $p = .19$, 6 males and 8 females were present in control condition and 1 male and 6 female were in the experimental condition), and ethnicity ($X^2 = .75$, $p = .861$, control condition included 10 Caucasian, 2 Asian, 1 Biracial, and 1 African-American student, while the experimental condition included 5 Caucasians, 1 Asian, and 1 Biracial student).

Table 1.

Participant's Frequency of Demographic Information

Frequencies			
Frequencies of Age			
Levels	Counts	% of Total	Cumulative %
18	3	14.3 %	14.3 %
19	4	19.0 %	33.3 %
20	5	23.8 %	57.1 %
21	5	23.8 %	81.0 %
22	3	14.3 %	95.2 %
25	1	4.8 %	100.0 %
Frequencies of Year			
Levels	Counts	% of Total	Cumulative %
1	5	23.8 %	23.8 %
2	3	14.3 %	38.1 %
3	6	28.6 %	66.7 %
4	7	33.3 %	100.0 %
Frequencies of Gender			
Levels	Counts	% of Total	Cumulative %
1 = male	7	33.3 %	33.3 %
2 = female	14	66.7 %	100.0 %
Frequencies of ethnicity			
Levels	Counts	% of Total	Cumulative %
1 = Caucasian	15	71.4 %	71.4 %
2 = Asian	3	14.3 %	85.7 %
3 = Biracial	2	9.5 %	95.2 %
4 = African American	1	4.8 %	100.0 %

After filling out the demographic information, participants are asked to measure their anxiety levels during their tests through STAI (State-Trait Anxiety Inventory). Though not

related to our primary interest, measuring anxiety during performances seems helpful as many literatures, such as Mattarella-Micke et al. (2011), suggested that anxiety plays a big role on how stimulus affects students' academic performances.

STAI was developed by Spielberger (2010), inspired by the idea that there should be a distinction between anxiety as a state and anxiety as a personality trait, as someone can be calm as an anxious person at a certain state and vice versa. Because the current study's focus is not interested in whether someone tends to be anxious, but rather whether they are anxious during the time of performing the task, STAI was the appropriate self-reported measure. STAI consists of lists of certain states of feelings (such as, pleasant, calm, secure, and tense etc.) in which participants can rate, on a scale of 1-4, how much they resonated with those states during the time of the task. All negative states were added with reverse-scored positive states to result in a state anxiety level for the exams. Participants were asked to complete two separate STAI for tests of each difficulty.

An important part of the current study's hypothesis discusses whether extraversion affects the confederate's impact on participants' test taking ability. In order to measure a complicated part of a human being, such as personality, the best case scenario is to have abundant time and resources to explore and understand it individually in depth. However, given the need for a simple, quick, and effective way of measuring personality, this study used the TIPI (Ten-Item Personality Inventory) to measure extraversion. The TIPI shows acceptable ratings in terms of test-retest reliability and convergent and discriminant validity (Gosling, Rentfrow, & Swann, 2003).

The TIPI includes a total of 10 lists of characteristics. Those who take the inventory are asked to rate each characteristic set on a scale of 1–7, in terms of how much they see themselves as the given characteristic. Two sets of characteristics represent each personality from the Big Five Inventory : Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness to Experience (Gosling, Rentfrow,& Swann, 2003). The current study only focuses on extraversion part of the personality. However, the current study asked all participants to rate all 10 characteristics instead of just the two that signal extraversion. This is because (a) there were no suggestions that the altered test or only taking a specific part of the test of interest will maintain its same validity and reliability as its full test, and (b) the TIPI was already a very short inventory that was able to be finished in a short amount of time.

Materials collected from the last two sections of the survey by all participants (participants' perceptions of test 1 and of test 2) are not used to discuss primary findings. Rather it is present to serve as a partial manipulation check as well as a light inquisition of whether presence of anxiety plays a role in the impact of the confederate, if any. The sections for participants' perception of the tests focused on measuring the validity of the participants' task performance. For example, the section asked whether the participant agreed or disagreed in a 7 point scale of statements such as : *I have performed this test to the best of my ability* or *Performing well on this test means a lot to me* . If the majority of the participants did not perform to their best ability (above 4), and hence will not illustrate how they would react to an actual test environment, the validity of the measures of task performance in current study should be questioned. Lastly, the section double checks whether the difficulty of the tests were perceived as intended. Because part of the current study's hypothesis is interested in the influence of difficulty

of the task, under no circumstances can a participant feel that the first test was harder than the second.

Lastly, a section was added for participants in the experimental condition, alongside with one verbal questionnaire to manipulation check for the presence and effect of the confederate. In a seven scale, questions asked how they feel agree or disagree with given statements to check if the participants perceived the confederate as intended (confident, well-performing and intelligent.) as well as whether the action of *handing the test early* of the confederate had any direct impact. Some statements included in the section are as follows : *The participant who did the next to me was confident; when the participant next to me handed in their test, I worked harder and faster.* It must be known whether the majority of the participants perceived the confederate as intended to assure the validity of the current study. Furthermore, the researcher asked every participant in the experimental condition, whether they believed the confederate was a participant. Unless it was an absolute no, the participant passed this manipulation check. Upon completion of the manipulation check, all participants were thoroughly debriefed.

Results

Of the total of 21 participants, 7 were placed in experimental condition and 14 were placed in the control condition. All participants in the experimental condition passed the verbal manipulation check by confirming that they did not recognize that the confederate was not another participant.

To test the hypothesis of whether having a confident peer affects performance in terms of time or test scores, two-way ANOVAs were conducted for Difficulty x Condition predicting test time and test score. To test whether there was an interaction between Condition and Difficulty

predicting test time, a repeated measure ANOVA showed no significant interaction, $F(1, 19) = 1.23, p = .28$. To test whether there was an interaction between Condition and Difficulty

predicting test score, a repeated measure ANOVA showed no significant interaction.

To test the main effects, a one-way ANOVA was conducted statistically predicting score or time by Condition. For Test 1 (easy), test score was not statistically predicted by condition, $F(1, 19) = 1.24, p = .28$, nor was test time, $F(1, 19) = 1.95, p = .66$. For Test 2 (difficult), test score was not statistically predicted by condition, $F(1, 19) = .26, p = .62$, but test time was, $F(1, 19) = 4.50, p = .047$.

Estimated Marginal Means

Condition * Difficulty

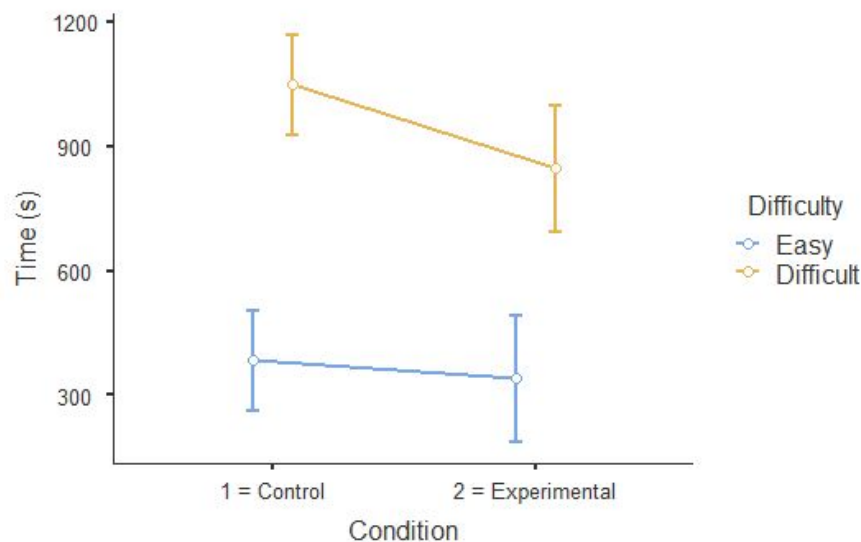


Figure 1. Difference in time (seconds) taken to finish the easy (first) and difficult (second) task between conditions. In terms of mean average, participants in the experimental condition performed quicker than participants in the control condition. Yet, the difference is only statistically significant in the difficult test and not in the easy test. There is not significant interaction between Condition x difficulty for time.

Table 2.

ANOVA test for interaction between Group x Difficulty on Test Score

Repeated Measures ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Difficulty	104.07	1	104.07	68.76	< .001
Difficulty * Condition	1.86	1	1.86	1.23	0.281
Residual	28.76	19	1.51		

Note. Type 3 Sums of Squares

[4]

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Condition	0.00298	1	0.00298	9.81e-4	0.975
Residual	57.61607	19	3.03242		

Note. Type 3 Sums of Squares

Table 3.

ANOVA test for interaction between Group x Difficulty on Test Times

Repeated Measures ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Difficulty	3.20e+6	1	3.20e+6	53.979	< .001
Difficulty * Condition	58196	1	58196	0.982	0.334
Residual	1.13e+6	19	59259		

Note. Type 3 Sums of Squares

[4]

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Condition	140402	1	140402	4.92	0.039
Residual	542652	19	28561		

Note. Type 3 Sums of Squares

To test the hypothesis of whether gender impacts the prediction that test performance will moderate the relationship between difficulty and condition, the study focused on the three way interaction between Condition, Difficulty, and Gender on performance first, with Difficulty as

repeated measures factor, and between subject factors of Condition and Gender. For the difference in time taken to complete the task, there was no significant result, $F(1, 17) = .07, p = .80$; for difference in test scores, there was also no significant result, $F(1, 17) = 2.38, p = .14$.

Table 4.

ANOVA test for 3-way interaction between Group x Difficulty x Gender on Test Score

Repeated Measures ANOVA

Within Subjects Effects					
	Sum of Squares	df	Mean Square	F	p
Difficulty	41.6430	1	41.6430	29.2370	< .001
Difficulty * Condition	0.0930	1	0.0930	0.0653	0.801
Difficulty * Gender	3.8573	1	3.8573	2.7082	0.118
Difficulty * Condition * Gender	3.3930	1	3.3930	2.3822	0.141
Residual	24.2135	17	1.4243		

Note. Type 3 Sums of Squares

[4]

Between Subjects Effects					
	Sum of Squares	df	Mean Square	F	p
Condition	0.976	1	0.976	0.307	0.587
Gender	3.483	1	3.483	1.094	0.310
Condition * Gender	1.343	1	1.343	0.422	0.525
Residual	54.130	17	3.184		

Note. Type 3 Sums of Squares

Table 5.

ANOVA test for 3-way interaction between Group x Difficulty x Gender on Test Time

Repeated Measures ANOVA

Within Subjects Effects					
	Sum of Squares	df	Mean Square	F	p
Difficulty	1.85e+6	1	1.85e+6	28.4981	< .001
Difficulty * Condition	31388	1	31388	0.4842	0.496
Difficulty * Gender	5194	1	5194	0.0801	0.781
Difficulty * Condition * Gender	4382	1	4382	0.0676	0.798
Residual	1.10e+6	17	64824		

Note. Type 3 Sums of Squares

[4]

Between Subjects Effects					
	Sum of Squares	df	Mean Square	F	p
Condition	109600	1	109600	3.5356	0.077
Gender	12928	1	12928	0.4171	0.527
Condition * Gender	696	1	696	0.0224	0.883
Residual	526988	17	30999		

Note. Type 3 Sums of Squares

Lastly, to test the hypothesis that extraversion impacts the prediction of test performance by condition, the three way interaction between Condition, Difficulty, and Extraversion was followed up with repeated measures ANOVA, with factors of Condition, Difficulty, and Extraversion. Extraversion scores on the TIPI were used to divide participants into three groups . The repeated measures ANOVA showed a significant three-way interaction when predicting test performance in terms of score, $F(2, 15) = 4.50, p = .029$, but not in terms of time, $F(2,15) = 1.25, p = .32$.

Table 6.

ANOVA test for 3 way interaction between Group x Difficulty x Extraversion on Test Score

Repeated Measures ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Difficulty	62.7011	1	62.7011	53.9107	< .001
Difficulty * Condition	0.0828	1	0.0828	0.0712	0.793
Difficulty * Extraversion labeled	2.2565	2	1.1283	0.9701	0.402
Difficulty * Condition * Extraversion labeled	10.4708	2	5.2354	4.5014	0.029
Residual	17.4458	15	1.1631		

Note. Type 3 Sums of Squares

[4]

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Condition	1.37	1	1.37	0.508	0.487
Extraversion labeled	3.54	2	1.77	0.655	0.534
Condition * Extraversion labeled	16.23	2	8.11	3.006	0.080
Residual	40.48	15	2.70		

Note. Type 3 Sums of Squares

Table 7.

ANOVA test for 3 way interaction between Group x Difficulty x Extraversion on Test Time

Repeated Measures ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Difficulty	1.74e+6	1	1.74e+6	29.432	< .001
Difficulty * Condition	10298	1	10298	0.174	0.682
Difficulty * Extraversion labeled	133075	2	66538	1.125	0.351
Difficulty * Condition * Extraversion labeled	147280	2	73640	1.245	0.316
Residual	887432	15	59162		

Note. Type 3 Sums of Squares

[4]

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Condition	29052	1	29052	1.30	0.272
Extraversion labeled	141543	2	70771	3.17	0.071
Condition * Extraversion labeled	65930	2	32965	1.48	0.260
Residual	335123	15	22342		

Note. Type 3 Sums of Squares

Though not included in the hypothesis, STAI for each participant were measured to see if there were any differences between state anxiety levels two conditions. Two separate Independent T-tests were conducted for state anxiety levels between participants in the control group and experimental group. Difference in state-anxiety level was present between participants in the control group and participants in the experimental group for the first (easy) exam ($t[19] = -3.43, p = 0.003, M_{\text{control}} = 36.6, M_{\text{experimental}} = 50, \text{Mean difference} = 13.4$), as well as the second (difficult) exam ($t[19] = -2.46, p = 0.0024, M_{\text{control}} = 42.8, M_{\text{experimental}} = 52.7, \text{Mean difference} = 10.9$)

Table 8.

Difference in State-Anxiety levels between Conditions in Varying Difficulties.

Independent Samples T-Test

Independent Samples T-Test		statistic	df	p	Mean difference	SE difference	Cohen's d
State Anxiety Level Test 2	Student's t	-2.23 *	18.0	0.039	-10.7	4.81	-1.09
State Anxiety Level Test 1	Student's t	-3.09	18.0	0.006	-13.0	4.21	-1.51

* Levene's test is significant ($p < .05$), suggesting a violation of the assumption of equal variances

Discussion

The current study mainly hypothesized that, in the presence of a seemingly confident, and extremely fast test taking peer, participants will, in general, show improved performances during easier tasks but impaired performances during difficult tasks in terms of accuracy, and overall improvement of the task in terms of time. The data found only supported such hypotheses partially. Participants showed significant improvement in terms of time, that is, it took the participants significantly less time to finish the difficult test, when they were with a confident peer who finished the test early. However, unlike the initial prediction, such improvement in time was not reflected in the easier test. Moreover, unlike the study's hypothesis, no difference of test performance in terms of accuracy was found in either of the difficulties.

The current study further was interested in finding whether other personal variables such as gender or personality (specifically extraversion) would have an impact on the condition's effect of participant's test performance, if any. Interestingly, the effect of condition and difficulty on participants were affected by extraversion (introverted, neutral, extroverted) but not by sex (male, female).

The measure of participant's anxiety level was not part of the main hypothesis, nor was part of the pre-registration of the current study. However, as many literatures discussed earlier, such as that of Mattarella-Micke et al.(2011) or Benny and Banks (2015), anxiety can potentially play a big role in a student's academic performance during a high-pressure situation.

Interestingly, all participants in the experimental condition showed significantly more levels of state-anxiety than those who were in the control condition, even though there were no significant differences in their actual academic performances. Though not concerned with our hypothesis, this is an incredibly interesting finding. This suggests at least two things: (a) the mere presence of the unfamiliar confident confederate is enough to drastically increase a student's anxiety level; (b) anxiety alone is not enough of a variable to affect a student's task performance ability. This can be used to further

There are many potential limitations with the participants. First and foremost, it is unequivocal that not only the sample size is small, but also, disproportioned, where the control condition had about as twice as many participants than as the experimental condition. With limited time and resources as the nature of being a year-long project, the current study was designed to be a pilot study for future studies refurbish and explore. Regardless, the researcher ambitiously aimed for 40-80 participants with each condition having equal amounts of participants. Such target sample size was initially lowered to 30 (15 participants in each condition) due to the unexpectedly low relative interest. This change also was realistic as the confederate was also a current Bard undergrad student and only could help out with so many data collection without getting paid; the current study preferred not to use more than one confederate as any difference in the confederates may lead to an uncontrolled variable.

Furthermore, the sample size was again lower than anticipated, as data collection was forced to be closed early due to the COVID19 pandemic. Remaining scheduled participants either left campus before they could participate in the study, or no longer felt comfortable participating. Getting more participants or data collection was forced to cease mid-March, which was right before the weekend the remaining seven participants were scheduled to participate in the study.

The current test used the subject math as the measure test performance. Therefore it is important that both groups have similar levels of math skills. When the study's method was being designed, the researcher hoped that the big sample size (initially planned to be 80) would be able to control for exceptional outliers and will control for variables math skills and demographic information. However with a drastic cut in sample size, current study did not control for math skills or general demographic information between two groups. Fortunately, participants were not statistically different between conditions from each other in terms of demographic information except for class year. Notably, there were no seniors or sophomores present in the experimental conditions whereas the control group had 7 seniors and 3 sophomores. However, there still may be differences in innate math skills and academic caliber between the participants in each group which increases the potential for errors.

Regardless of the attempt, the current study failed to functionally perform random sampling. Even though the target sample was limited in a sense that it only aimed for students in one undergrad institution, the current study attempted to catch as diverse members possible through advertising in different platforms. Either physical or digital poster advertising for the study's participation was posted on Facebook's Bard Students group page, Instagram, as well as

physical copy posted throughout college campus. However, more than half the participants personally knew the researcher beforehand the study and was conveniently sampled to participate in the data collection. Because of this, the majority of the sample is even more limited to Bard undergrad students who know the researcher and does not well represent the general population.

Critically, the current study could not functionally apply random assignment due to the limited availability of the confederate. Initially, a random number generator in google (min 1,max 10) was used to randomly assign participants to either condition. Those who were given odd numbers were assigned to the control condition and those who were given even numbers were assigned to the experimental condition. With the confederate's very limited availability, the schedule for participants who were randomly assigned to the experimental conditions and confederate did not always line up. Because the current study already was lacking volunteers, instead of waiting for participants who were randomly assigned to experimental conditions and had compatible availability to the confederate, the study opted to not perform random assignment. Participants who were available at the time when the confederate was as well were assigned to the experimental condition and the rest were assigned to the control condition.

Though not perfectly organized, the current study ambitiously challenged to find the connection between many important variables that seem to have some connection to task performances. I hope this study can ignite a further exploration in the relationship between test performances, anxiety, presence of others, difficulty, and pressure. A replication of this study with more robust sample size will be a good start. I hope to ultimately see where we can provide a solution for students to perform with the best of their ability in different scenarios.

References

Beilock, S. (2010). *Choke : What the Secrets of the Brain Reveal about Getting it Right When You have to*. Simon and Schuster

Benny, B. & Banks, J. (2015). Under pressure: An examination of the predictors of choking. *Journal of Individual Differences*, 36(2), 93-100.

<https://psycnet.apa.org/record/2015-16536-005>

Bradley, B. H., Klotz, A. C., Postlethwaite, B. E., & Brown, K. G. (2013). Ready to rumble: How team personality composition and task conflict interact to improve performance. *Journal of Applied Psychology*, 98(2), 385–392. <https://doi.org/10.1037/a0029845>

Breevaart, K., & Zacher, H. (2019). Daily selection, optimization, and compensation strategy use and innovative performance: The role of job autonomy and time pressure. *Journal of Personnel Psychology*, 18(2), 71–83. <https://doi.org/10.1027/1866-5888/a000224>

Carsten, T., Hoofs, V., Boehler, C. N., & Krebs, R. M. (2019). Are losses more effective than rewards in improving performance in a cognitive task? *Motivation Science*, 5(3), 257–268. <https://doi.org/10.1037/mot0000117.supp> (Supplemental)

Chemers, M. M., Hu, L., & Garcia, B. F. (2001). Academic self-efficacy and first year college student performance and adjustment. *Journal of Educational Psychology*, 93(1), 55–64. <https://doi.org/10.1037/0022-0663.93.1.55>

Deci, E. L., & Ryan, R. M. (2012). Self-determination theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (p. 416–436). Sage Publications Ltd. <https://doi.org/10.4135/9781446249215.n21>

Deci, E. L. & Ryan, R. M. (1985). The General Causality Orientations Scale : Self Determination in Personality. *Journal of Research in Personality*, 19, 109-134.

https://selfdeterminationtheory.org/SDT/documents/1985_DeciRyan_GCOS.pdf

Doren, B., Gau, J. M., & Lindstrom, L. E. (2012). The Relationship Between Parent Expectations and Postschool Outcomes of Adolescents With Disabilities. *Exceptional Children*, 79(1), 7–23. <https://doi.org/10.1177/001440291207900101>

Esposito, G., Karterud, S., Freda, F. (2019). Mentalizing underachievement in group counseling: Analyzing the relationship between members' reflective functioning and counselors' interventions. *Psychological Services*. 1541-1559.

<https://www.ncbi.nlm.nih.gov/pubmed/30932505>

Gelso, C., Osterhouse, R., Bodden, J. (1971). Nonintellective factors in improvement during an educational skills course. *Journal of Counseling Psychology*, v18(5), 503-504. <https://psycnet.apa.org/record/1972-03784-001>

Gosling, S., Rentfrow, P., Swann Jr., W. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Psychology*, 37, 504-528.

<http://gosling.psy.utexas.edu/wp-content/uploads/2014/09/JRP-03-tipi.pdf>

Helbling, L. A., Tomasik, M. J., & Moser, U. (2019). Long-term trajectories of academic performance in the context of social disparities: *Longitudinal findings from Switzerland*. *Journal of Educational Psychology*, 111(7), 1284–1299. <https://doi.org/10.1037/edu0000341>

Kim, A. S. N., Shakory, S., Azad, A., Popovic, C., & Park, L. (2019). Understanding the impact of attendance and participation on academic achievement. *Scholarship of Teaching and Learning in Psychology*. <https://doi.org/10.1037/stl0000151>

Kolb, K. J., & Jussim, L. (1994). Teacher expectations and underachieving gifted children. *Roeper Review*, 17(1), 26. <https://doi.org/10.1080/02783199409553613>

Kravitz, D. A., & Martin, B. (1986). Ringelmann rediscovered: The original article. *Journal of Personality and Social Psychology*, 50(5), 936–941.
<https://doi.org/10.1037/0022-3514.50.5.936>

Kretschmann, J., Vock, M., Lüdtke, O., Jansen, M., & Gronostaj, A. (2019). Effects of grade retention on students' motivation: A longitudinal study over 3 years of secondary school. *Journal of Educational Psychology*, 111(8), 1432–1446. <https://doi.org/10.1037/edu0000353>

Mattarella-Micke, A., Mateo, J., Kozak, M., Foster, K., Beilock, S. L. (2011). Choke or thrive? The relation between salivary cortisol and math performance depends on individual difference in working memory and math anxiety. *Emotion*, 11(4), 1000-1005.
<https://www.ncbi.nlm.nih.gov/pubmed/21707166>

McDaniel, L. (1990). The Effects of Time Pressure and Audit Program Structure on Audit Performance. *Journal of Accounting Research*, 28(2), 267-285. doi:10.2307/2491150

Morris, L. W., & Liebert, R. M. (1970). Relationship of cognitive and emotional components of test anxiety to physiological arousal and academic performance. *Journal of Consulting and Clinical Psychology*, 35(3), 332–337. <https://doi.org/10.1037/h0030132>

Prem, R., Paškvan, M., Kubicek, B., & Korunka, C. (2018). Exploring the ambivalence of time pressure in daily working life. *International Journal of Stress Management*, 25(1), 35–43. <https://doi.org/10.1037/str0000044>

Rosenthal, R., Jacobson, L. Pygmalion in the classroom. *Urban Rev* 3, 16–20 (1968).
<https://doi.org/10.1007/BF02322211>

Rubie-Davies, C. M., Peterson, E., Irving, E., Widdowson, D., & Dixon, R. (2010). Expectations of achievement Student, teacher and parent perceptions. *Research in Education*, 83(1), 36–53. <https://doi.org/10.7227/RIE.83.4>

Rubovits, P. C., & Maehr, M. L. (1971). Pygmalion analyzed: Toward an explanation of the Rosenthal-Jacobson findings. *Journal of Personality and Social Psychology*, 19(2), 197–203. <https://doi.org/10.1037/h0031526>

Seashore, C. E. (1899). The Dynamogenic Factors in Pacemaking and Competition. *Psychological Review*, 6(3), 336. <https://doi.org/10.1037/h0069301>

Schaeffer, M. W., Rozek, C. S., Berkowitz, T., Levine, S. C., Beilock, S. L. (2018). Disassociating the relation between parents' math anxiety and children's math achievement : Long-term effects of a math app intervention. *Journal of Experimental Psychology*, v147(12), 1782-1790. <https://psycnet.apa.org/doiLanding?doi=10.1037%2Fxxg0000490>

Song, J., Bong, M., Lee, K., & Kim, S. (2015). Longitudinal investigation into the role of perceived social support in adolescents' academic motivation and achievement. *Journal of Educational Psychology*, 107(3), 821–841. <https://doi.org/10.1037/edu0000016>

Spielberger, C. D. (2010). State-Trait Anxiety Inventory. *The Corsini Encyclopedia of Psychology*. <https://doi.org/10.1002/9780470479216.corpsy0943>

Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology*, 67(1), 415–437.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1), 4–28. File

Tenney, E. R., Logg, J. M., & Moore, D. A. (2015). (Too) optimistic about optimism: The belief that optimism improves performance. *Journal of Personality and Social Psychology*, 108(3), 377–399. <https://doi.org/10.1037/pspa0000018.supp> (Supplemental)

Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *The American Journal of Psychology*, 9(4), 507-533.

Wilson, T., Linville, P. (1982). Improving the academic performance of college freshmen : Attribution therapy revisited. *Journal of Personality and Social Psychology*, v42(2). 367-376. <https://psycnet.apa.org/record/1982-26799-001>


Wirth, B. E., & Carbon, C.-C. (2017). An easy game for frauds? Effects of professional experience and time pressure on passport-matching performance. *Journal of Experimental Psychology: Applied*, 23(2), 138–157. <https://doi.org/10.1037/xap0000114.supp> (Supplemental)

Wright, S. M. (2020, February 6). Flipping the Script: Utilizing a Modified Harvard Case Study Model in Undergraduate Life Span Development Classes. *Scholarship of Teaching and Learning in Psychology*. Advance online publication. <http://dx.doi.org/10.1037/stl0000183>

Zajonc, R. B., Heingartner, A., & Herman, E. M. (1969). Social enhancement and impairment of performance in the cockroach. *Journal of Personality and Social Psychology*, 13(2), 83–92. <https://doi.org/10.1037/h0028063>

Appendix

a. Preregistration

 <div style="margin-left: 20px;"> <h2 style="margin: 0;">OSL/PT Preregistration Template</h2> </div>	
<p>Investigator's Name and Affiliation (leave blank if this is an anonymous preregistration)</p> <div style="border: 1px solid black; padding: 2px; margin-bottom: 10px;">Woanjun Lee</div> <p>Names and Affiliations of Collaborators (leave blank if this is an anonymous preregistration)</p> <div style="border: 1px solid black; padding: 2px; min-height: 50px;">Advised by Prof. Justin Dainer-Best</div>	<p>Date of Preregistration February 1, 2020</p> <p>IRB Status</p> <p><input type="radio"/> IRB Review Not Necessary</p> <p><input type="radio"/> Not Submitted Yet</p> <p><input type="radio"/> Submitted</p> <p><input checked="" type="radio"/> Approval Received, Date: Dec. 16th 2019</p>
<p>Study Title</p> <div style="border: 1px solid black; padding: 2px; min-height: 20px;">Friend or Foe? Confident Peer's Unexpected Impact on Test Taking Ability</div>	
<p>VARIABLES</p> <p>What are your independent / grouping / predictor variables (including mediators and moderators) ? Explain how you operationalize each variable.</p> <div style="border: 1px solid black; padding: 5px; min-height: 60px;"> <p>Main independent variable is the presence of a confederate taking the exam alongside with the participant. Though each participant in both experimental group and control group will be taking the same exam, participants in control group will be taking the test alone, while the participants in the experimental group will be taking the test while the confederate is also taking the test next to them. Though confederates are asked to take the test in a very confident manner and finish it significantly earlier than the participant, the study is designed so that the participant will not directly interact with the confederate during the test.</p> </div> <p>What are your dependent / outcome variables? Explain how you operationalize each variable.</p> <div style="border: 1px solid black; padding: 5px; min-height: 60px;"> <p>Main dependent variable is test performance in terms of accuracy and speed. Accuracy will be measured by how many questions students got the question correctly and speed will be measured on how fast it took for participants to finish the test. To note, participants are given no time limit to finish their task.</p> </div> <p>List any exploratory variables. These are variables that you included in your study, but are not central to your main predictions.</p> <div style="border: 1px solid black; padding: 5px; min-height: 60px;"> <p>Difficulty of the Test : one is more challenging.</p> <p>Sex : demographic information of participants will be collected upon their consent; current study believes that any finding of the main hypothesis may be different depending on the sex of the participant.</p> <p>Personality: A brief personality test will be used to post task to reflect very basic level of personality traits; current study</p> </div>	

Did you create new, or modify existing, variables for this study? (select all that apply)

- ☒ Some, or all, variables have been used in prior, published research, and no modifications were made
- ☐ Some variables were modified from their original form
- ☒ Some variables were created for this study

If you indicated above that 'Some variables were modified,' describe how you modified existing variables here:

If you indicated above that 'Some variables were created for this study,' list and describe the variables that you created for this study:

-Peer Presence : whether or not confident task performer is present alongside of participant
-test difficulty
-Math test:

HYPOTHESES

What are your primary study hypotheses / research questions?

This study hypothesizes that in the presence of a confident, fast test taking peer, the participant will show improved performances during easier tasks but impaired performances during difficult tasks

Do you have any exploratory hypotheses / research questions? If so, describe them below:

-Does Gender have an impact on the effect of peer presence?
-Does personality have an impact on the effect of peer presence?
-Does Task difficulty have an impact on the effect of peer presence?

At the time of this preregistration, describe the status of data collection:

- ☐ No new data collection is required for this project (e.g., meta-analysis)
☒ Data collection has not started for this study
☐ Data collection is in progress
☐ Data collection is complete
☐ Other:

If you selected 'Other' to describe the status of data collection, please explain here:

If you indicated above that data collection is 'complete' or 'in progress,' have you (or anyone else) already conducted any statistical analyses?

- ☐ No data analyses have been performed
☐ Some preliminary analyses have been performed, but not those relevant to the primary or exploratory study hypotheses described above (e.g., you calculated descriptive statistics)
☐ Some, or all, analyses of the primary or exploratory hypotheses have been performed

If you selected 'Some preliminary analyses have been performed' describe the analyses you have already conducted:

If you selected 'Some, or all, analyses of the primary or exploratory hypotheses have been performed,' you should stop completing this form. Pre-registration of hypotheses MUST occur before you have analyzed your data.

SAMPLING

What is your target sample size?

How was your target sample size determined? (check all that apply)

- ☐ Power analysis
☐ Target sample size based on convention / past research
☒ Target sample size based on constraints / convenience (e.g., size of subject pool, available money to pay participants, access to participants)
☐ Other:

How will you determine when to stop collecting data (i.e., your stopping rule)?

- ☒ When the target sample size is reached
☐ A particular amount of time has passed (e.g., the end of the semester)
☒ Other (describe below)

If you selected 'Other' for your stopping rule, please explain here:

Target sample size is reached OR March 15, 2020

RESEARCH DESIGN

What type of research design are you using?

- ☒ Experiment
☐ Quasi-experiment
☐ Correlational Study
☐ Other:

If you selected 'Other' for your research design, please explain here:

EXPERIMENTAL DESIGNS ONLY

If you are conducting an experiment, what is the nature of the manipulation?

- ☒ between-participants
☐ within-participants
☐ mixed (at least one between and one within factor)

What are the total number conditions in your study? (e.g., a 2 x 2 design has 4 total conditions):

Will the experimenters be aware of the condition to which a particular participant has been assigned?

- ☒ Yes, the experimenter will be aware of the condition to which a participant has been assigned
☒ No, the experimenter will be blind to condition

Will participants be randomly assigned to condition?

- ☒ Yes
☒ No (describe below)

If you selected 'No' for how you will assign participants to condition, please explain here:

Random but with caveats for scheduling.

If you are predicting an interaction (in your hypotheses), describe the nature of that interaction below:

Participants are likely to perform worse in presence of confident peer, especially when the task is difficult, or if the participant is a female, or if the participant has anxious or introverted personalities.

DATA ANALYSIS PLAN

What will be your criterion for determining statistical significance?

- ☒ $p < .05$ ☐ $p < .01$ ☐ $p < .005$
☐ Other:

Will your tests of significance be:

- ☒ One-tailed
☒ Two-tailed
☐ A combination of one- and two-tailed tests

If you indicated that some tests of significance will be one-tailed, describe the hypothesis and predicted direction of the effect or association below:

Will you exclude participants from data analysis based on any of the reasons listed below?

☐ Failed attention check ☒ Failed manipulation check ☒ Missing data

Describe any additional exclusion criteria here:

What criterion (if any) will you use to determine whether a participant is an outlier?

☒ Greater than 3 standard deviations from the mean

☐ Other:

Which statistical tests will you use to conduct your data analyses? (check all that apply)

☒ ANOVA ☐ Correlation ☐ t-test ☐ Chi-square
☐ Regression ☐ Other/Additional

If you selected 'Other/Additional' for the statistical test above, describe the analyses you will conduct here:

If relevant, describe what types of follow-up tests will you perform (e.g., Tukey post-hoc; simple main effects). If you will conduct planned comparisons, explain the nature of those comparisons below:

For the analyses listed above, will you include any covariates or control variables? If so, describe them below and provide a justification:

This preregistration template was created by Kevin P. McIntyre, kmcintyr@trinity.edu, Trinity University, and Benjamin Le, ble@haverford.edu, Haverford College. For more information, visit www.openstatslab.com and www.projecttier.org or follow us @openstatslab @Project_TIER

© 2017 v1.0 This material may be used for noncommercial educational and research purposes only.

a. First test

For the following inequalities, solve for x

$$x + 5 = 3$$

$$4(x - 1) = 3(x + 1)$$

$$2x + 3 = x$$

$$-12x = 144$$

$$5[x + 3(1 - x) + 2] = 7.$$

$$7x + 2 = 65$$

$$5x - 25 = 0$$

b. Second Test

(1) Twenty percent of the sweaters in a store are white/ Of the remaining sweaters, 40 percent are brown, and the rest are blue. If there are 200 sweaters in the store, then how many more blue sweaters than white sweaters are in the store?

$$(2) (4^{13} - 4^{12})/4^{11} = ?$$

What is ?

- a) 0
- b) 1
- c) 4
- d) 12
- e) 16

(3) If $a = (27)(3^{-2})$ and $x = (6)(3^{-1})$, then which of the following is equivalent to $(12)(3^{-x})$ multiplied by $(15)(2^{-a})$?

hint:

$$b^{-g} = 1/b^g$$

$$\text{ie : } 4^{-2} = 1/4^2$$

- a) $5(-2245)(320)$
- b) $\%$
- c) $5/2$
- d) $5(24)(38)$
- e) $5(2245)(320)$

(4) Jill has received 8 of her 12 evaluation scores. So far Jill's average (arithmetic mean) is 3.75 out of a possible 5. If Jill needs an average of 4.0 points to get a promotion, which list of scores qualifies Jill to receive her promotion

Indicate ALL such lists.

- (a) 3.0, 3.5, 4.75, 4.75
- (b) 3.5, 4.75, 4.75, 5.0
- (c) 3.25, 4.5, 4.75, 5.0
- (d) 3.75, 4.5, 4.75, 5.0

(5) If the probability of selecting, without replacement, 2 red marbles from a bag containing only red and blue marbles is $\frac{3}{55}$ and there are 3 red marbles in the bag, what is the total number of marbles in the bag?

- (a) 10
- (b) 11
- (c) 55
- (d) 110
- (e) 165

(6) All first-year students at Blue State University must take calculus, English composition, or both. If half of the 2,400 first-year students at Blue State University take calculus and half do not, and one-third of those who take calculus also take English composition, how many students take English composition?

- (a) 400
- (b) 800
- (c) 1,200

(d) 1,600

(e) 2,000

(7) A certain punch is created by mixing two parts soda and three parts ice cream. The soda is 4 parts sugar, 5 parts citric acid, and 11 parts other ingredients. The ice cream is 3 parts sugar, 2 parts citric acid, and 15 parts other ingredients.

Given the following,

Quantity A : Parts sugar in the punch

Quantity B : Parts Citric Acid in the punch

Select one of the four answer choices below.

(a) Quantity A is greater than Quantity B

(b) Quantity B is greater than Quantity A

(c) The two quantities are equal

(d) The relationship cannot be determined from the information given

D. Answer Keys for Tests

First Test

1. -2
2. 7
3. -3
4. -12
5. 1.8
6. 9
7. 5

Second Test

1. 56
2. D
3. C
4. B&D
5. B
6. D
7. A

E. Post Task Survey

Post Task Survey

Age :

Year at Bard :

Sex:

Ethnicity :

A) On the next page, there will be a number of personality traits that may or may not apply to you.

You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.

Please Check what is most applicable : I see myself as.....

	Disagree Strongly	Disagree Moderately	Disagree a Little	Moderate	Agree a Little	Agree Moderately	Agree Strongly
Extraverted, enthusiastic.							
Critical, quarrelsome.							
Dependable, self-disciplined.							

[illegible]

B) Please recall when you were taking the **First Exam : How did you feel?**

	Not at all	somewhat	moderately	Almost always
pleasant				
Calm				
Secure				
Tense				
Strained				
At ease				
upset				
Worried over possible misfortunes				
satisfied				

frightened				
comfortable				
self-confident				
nervous				
jittery				
indecisive				
relaxed				
content				
worried				
confused				
steady				

C) Please recall when you were taking the **Second Exam** : How did you feel?

	Not at all	somewhat	moderately	Almost always
--	-------------------	-----------------	-------------------	----------------------

pleasant				
Calm				
Secure				
Tense				
Strained				
At ease				
upset				
Worried over possible misfortunes				
satisfied				
frightened				
comfortable				
self-confident				

nervous				
jittery				
indecisive				
relaxed				
content				
worried				
confused				
steady				

D) For your **First Exam**, how much do you agree with the following statements?

(1 being strongly disagree, 4 being neither agree or disagree, 7 being strongly agree)

	1	2	3	4	5	6	7
I have performed this test to the best of my ability							

Performing well on this test means a lot to me							
I have performed well in the First Test							
The First Test was easy							

E) For your **Second Exam, how much do you agree with the following statements?**

(1 being strongly disagree, 4 being neither agree or disagree, 7 being strongly agree)

	1	2	3	4	5	6	7
I have performed this test to the best of my ability							
Performing well on this test means a lot to me							
I have performed well in the Second Exam							
The Second Exam was easy							

F. Post-task Survey Extra section for Experimental Conditions

F) Please tell us how much you agree with the following statement regarding the

First Exam (1 being strongly disagree, 4 being moderate, 7 being strongly agree)

	1	2	3	4	5	6	7
The participant who did the task next to me was confident							
The participant who did the task next to me probably performed good at their task							
The participant who did the task next to me probably performed better than me							
The participant who did the task next to me seems intelligent							
When the participant next to me handed in their test, I worked harder and faster							

When the participant next to me handed in their test, it made me uncomfortable							
---	--	--	--	--	--	--	--